Ultra-rapid nanopore genome sequencing in a critical care setting

Short Author List:

John E. Gorzynski, D.V.M., Ph.D., Sneha D. Goenka, M.Tech., Euan A. Ashley, M.B, Ch.B., D.Phil.*, and others.

Full Author List:

John E. Gorzynski, D.V.M., Ph.D., Sneha D. Goenka, M.Tech., Kishwar Shafin, B.S., Tanner D. Jensen, B.S., Dianna G. Fisk, Ph.D., Megan E. Grove, M.S., Elizabeth Spiteri, Ph.D., Trevor Pesout, B.S., Jean Monlong, Ph.D., Gunjan Baid, M.S., Jonathan A Bernstein, M.D., Ph.D., Scott Ceresnak, M.D., Pi-Chuan Chang, Ph.D., Jeffrey W. Christle, Ph.D., Henry Chubb, M.B.B.S., Ph.D., Karen P. Dalton, B.S. , Kyla Dunn, M.S., Daniel R. Garalde, Ph.D., Joseph Guillory, M.S., Joshua Knowles, M.D., Ph.D. Alexey Kolesnikov, M.S., Michael Ma, M.D., Tia Moscarello, M.S., Maria Nattestad, Ph.D., Marco Perez, M.D., Maura RZ Ruzhnikov, M.D., Mehrzad Samadi, Ph.D., Ankit Sethia, Ph.D., Chris Wright, D.Phil., Courtney J. Wusthoff, M.D., Katherine Xiong, M.D., Tong Zhu, M.S., Miten Jain, Ph.D., Fritz J. Sedlazeck, Ph.D., Andrew Carroll, Ph.D., Benedict Paten, Ph.D., Euan A. Ashley, M.B, Ch.B., D.Phil.*

*Correspondence to euan@stanford.edu

1. Stanford University, Stanford, CA
   (J.E.G., S.D.G., E.S., T.D.J., J.A.B., S.C., J.W.C., H.C., K.P.D, J.K., M.M., M.P., M.R.Z.R., C.J.W., K.X., E.A.A.)
2. Stanford Health Care, Palo Alto, CA
   (D.G.F., M.E.G., T.M.)
3. Stanford Children's Health, Palo Alto, CA
   (K.D.)
4. University of California at Santa Cruz Genomics Institute, Santa Cruz, CA
   (K.S., T.P. J.M., M.J., B.P.)
5. Baylor College of Medicine, Houston, TX
   (F.J.S.)
6. Google Inc, 1600 Amphitheatre Parkway, Mountain View, CA
   (G.B., P.C., A.K., M.N., A.C.)
7. Oxford Nanopore Technologies, Oxford, UK
   (D.R.G., J.G., C.W.)
8. NVIDIA Corporation, 2788 San Tomas Expy, Santa Clara, CA
   (M.S., A.S., T.Z.)

Word Count: 613

To the editor: Rapid genetic diagnosis can guide clinical management, reduce cost, and improve prognosis in critically ill patients.[1–3] Although most critical care decisions must be made in hours, traditional testing requires weeks and rapid testing requires days. We demonstrate that genome nanopore sequencing can accurately and rapidly provide genetic diagnoses. Our workflow combines streamlined preparation of commercial nanopore sequencing, distributed cloud-basedbioinformatics, and a custom variant-prioritization approach (**Figure 1**). Between December 2020 and May 2021, we enrolled 12 patients at two hospitals in Stanford, California, who were, within the context of this small sample size, generally representative (with respect to race ethnicity and sex) of persons living in the United States and California (Table S1, S2). We obtained an initial genetic diagnosis in five of them (**Table S3**). The shortest time from arrival of blood in the lab to initial diagnosis was 7 hours and 18 minutes.

After diagnosing patient 1, we updated our bioinformatics framework to permit the transfer of terabytes of raw signal data to cloud storage in real time and distributed the data across multiple cloud computing machines to achieve near real-time base calling and alignment, reducing post-sequencing runtime (base-calling through alignment) by 93% to 34 minutes (the average of post-sequencing runtimes for patients 2 to 12) (**Table S5**).

Flow cells were washed and reused until exhaustion to reduce the sequencing cost per sample. Libraries were barcoded in cases 1-7 to prevent contaminations of patient samples. After patient 7 we benchmarked and adopted a barcode free method to rapidly generate genome sequences. [4] Removing barcoding accelerated sample preparation (by 37 minutes) to an average of 2.5 hours, enabling us to load a greater amount of patient DNA into each flow cell (from 155 ng to 333 ng), and increase pore occupancy (from 64% to 82%) (**Figure S1, S2, Table S4)**. Sequencing with 48 flow cells generated 200 gigabases of data. Our workflow generated 173 to 236 gigabases of data per genome, with a post-quality control alignment identity of 94% (**Figure S3**) and 46-64X autosomal coverage: ie, each base of each autosome was represented in 46 to 64 sequence reads (**Figure S4**). Half of the sequencing reads were 25 kilobases or longer (**Table S6).**

Small variants and structural variants were called after the reads were aligned to the GRCh37 human reference genome, generating a median of 4,490,490 SNVs and small INDELs.[5,6] Custom filtration and prioritization of variants with an ultra-rapid scoring system (**Figure S5**) substantially decreased the number of candidate variants for manual review to a median of 29 (16-53) for small variants and 22 (11-37) for structural variants (**Table S2**).

Each initial diagnosis was immediately reviewed by study physicians and, when available, bedside physicians and a consensus was reached as to whether the proposed variant represented the primary cause of the patient's presentation. Diagnostic variants were identified in five of twelve patients, aged 3 months to 57 years. The findings were immediately confirmed by a CLIA-certified laboratory and informed clinical management (including sympathectomy, heart transplant, screening, and changes in medication) for each of the five patients or their family members.

In one patient, a 3 month-old full-term infant who presented in status epilepticus, seizure semiologies included right gaze deviation with bilateral upper extremity clonic jerking and perioral myoclonictwitching. Interictal EEG revealed abundant predominantly posterior multifocal sharps. Brain MRI was normal. Eight hours and twenty-five minutes after enrollment, we identified a likely pathogenic heterozygous variant in *CSNK2B* (a variant and gene known to cause a neurodevelopmental disorder with early-onset epilepsy) and made a definitive diagnosis of *CSNK2B*-related disorder, or Poirier-Bienvenu neurodevelopmental syndrome. This result halted further planned diagnostic testing, facilitated disease-specific counseling and prognostication, and aided in management of epilepsy by providing insight about reported seizure types and treatment response to common anti-seizure medications. The results of a test by epilepsy gene panel (which excluded *CSNK2B*), ordered at the time of presentation, were received two weeks later: only multiple non-diagnostic variants of uncertain significance were reported.

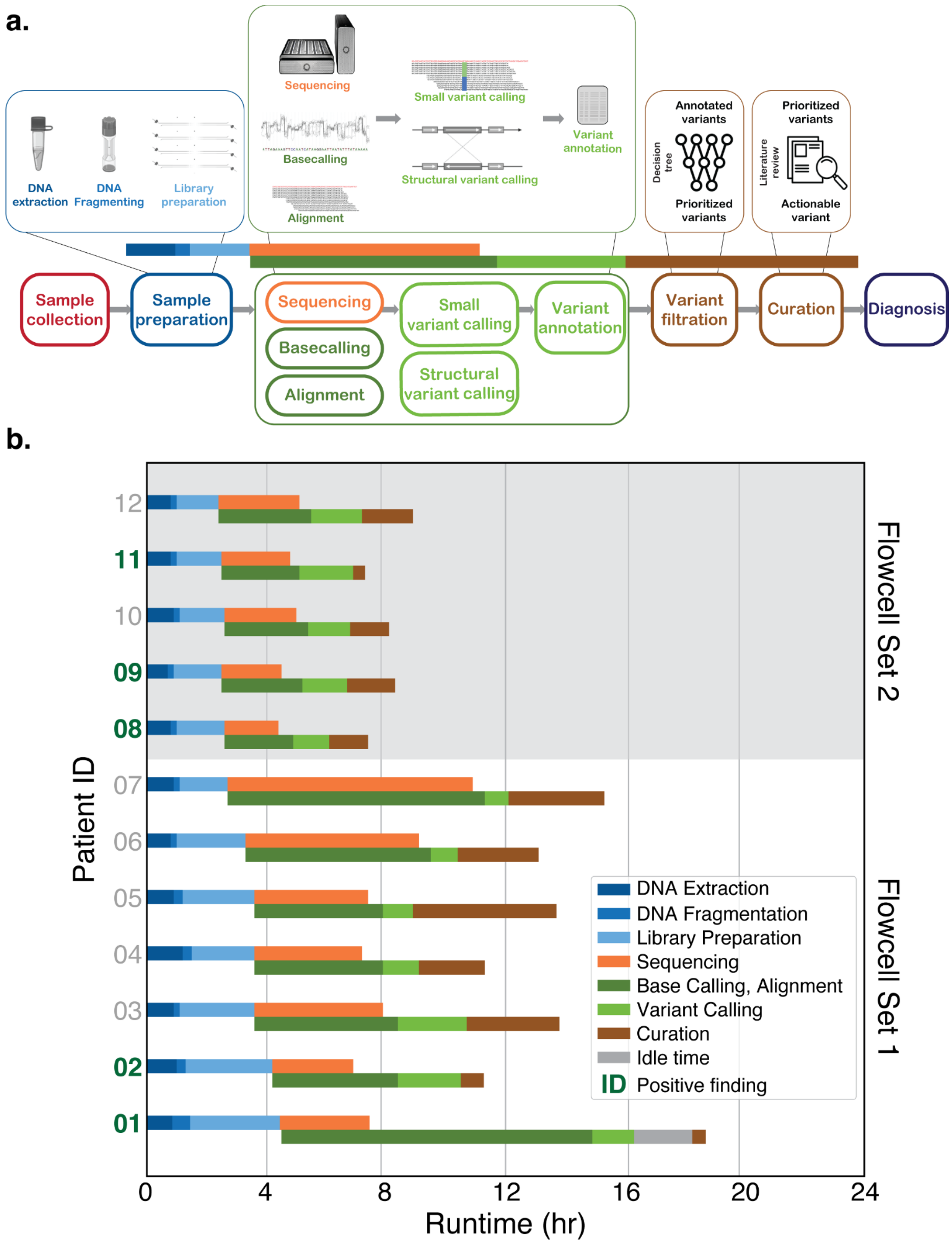


Figure 1. (a) The ultra-rapid genome sequencing pipeline. The schema depicts all processes from sample collection to a diagnosis. Vertically stacked processes are run in parallel. (b) The performance of the pipeline on twelve patients in two phases. Run-time of individual components are shown by corresponding color from panel (a). The fastest runtime was 7:18 hours (Patient 11) with a positive diagnosis.

References

1. Buchan JG, White S, Joshi R, Ashley EA. Rapid Genome Sequencing in the Critically Ill. Clin Chem 2019;65(6):723–6.

2. Dimmock D, Caylor S, Waldman B, et al. Project Baby Bear: Rapid precision care incorporating rWGS in 5 California children's hospitals demonstrates improved clinical outcomes and reduced costs of care. Am J Hum Genet [Internet] 2021;Available from: https://www.sciencedirect.com/science/article/pii/S0002929721001920

3. Priest JR, Ceresnak SR, Dewey FE, et al. Molecular diagnosis of long QT syndrome at 10 days of life by rapid whole genome sequencing. Heart Rhythm 2014;11(10):1707–13.

4. Sneha D. Goenka*, John E. Gorzynski*, Kishwar Shafin*, Dianna G. Fisk, Trevor Pesout , Jean Monlong , Tanner D. Jensen , Pi-Chuan Chang, Gunjan Baid, Jonathan A Bernstein, Jeffrey Christle , Karen P. Dalton, Daniel R. Garalde, Megan E. Grove, Joseph Guillory, Alexey Kolesnikov, Maria Nattestad, Maura R.Z. Ruzhnikov, Mehrzad Samadi, Ankit Sethia, Elizabeth Spiteri, Chris Wright, Katherine Xiong, Tong Zhu, Miten Jain, Fritz J. Sedlazeck, Andrew Carroll, Benedict Paten, Euan A. Ashley. Rapid distributed nanopore sequencing and clinical analysis of human whole genomes. In Press, Nature Biotechnology.

5. Shafin K, Pesout T, Chang P-C, et al. Haplotype-aware variant calling with PEPPER-Margin-DeepVariant enables high accuracy in nanopore long-reads. Nat Methods 2021;18(11):1322–32.

6. Sedlazeck FJ, Rescheneder P, Smolka M, et al. Accurate detection of complex structural variations using single-molecule sequencing. Nat Methods 2018;15(6):461–8.

In kind contributions from Oxford Nanopore, Google Inc., and Nvidia, and Stanford University unrestricted funds financially supported this study.

Supplementary Material for *Ultra-rapid whole genome nanopore sequencing in a critical care setting*.

John E. Gorzynski, D.V.M., Ph.D., Sneha D. Goenka, M.Tech., Kishwar Shafin, B.S., Tanner D. Jensen, B.S., Dianna G. Fisk, Ph.D., Megan E. Grove, M.S., Elizabeth Spiteri, Ph.D., Trevor Pesout, B.S., Jean Monlong, Ph.D., Gunjan Baid, M.S., Jonathan A Bernstein, M.D., Ph.D., Scott Ceresnak, M.D., Pi-Chuan Chang, Ph.D., Jeffrey W. Christle, Ph.D., Henry Chubb, M.B.B.S., Ph.D., Karen P. Dalton, B.S. , Kyla Dunn, M.S., Daniel R. Garalde, Ph.D., Joseph Guillory, M.S., Joshua Knowles, M.D., Ph.D. Alexey Kolesnikov, M.S., Michael Ma, M.D., Tia Moscarello, M.S., Maria Nattestad, Ph.D., Marco Perez, M.D., Maura RZ Ruzhnikov, M.D., Mehrzad Samadi, Ph.D., Ankit Sethia, Ph.D., Chris Wright, D.Phil., Courtney J. Wusthoff, M.D., Katherine Xiong, M.D., Tong Zhu, M.S., Miten Jain, Ph.D., Fritz J. Sedlazeck, Ph.D., Andrew Carroll, Ph.D., Benedict Paten, Ph.D., Euan A. Ashley, M.B, Ch.B., D.Phil.

" #$%&' (!

Patient recruitment
Enrollment was open to any critical care patient at Stanford hospitals (Stanford Health Care and Lucile Packard Children's Hospital) with a clinical presentation consistent with a genetic disease. Priority was given to patients where a rapidly identified genetic diagnosis would be clinically impactful for the patient or the patient's family. Recruitment of patients was loosely consecutive. While we did not screen all hospital in-patients our study relied on hospital clinicians to contact us in the event that a patient was a potential candidate. We acquired consent from adults directly and for minors, from parents or guardians according to Stanford IRB protocol 58559.

Sample collection
2cc of blood was collected either through sampling of an IV catheter or central line, or via venipuncture and stored in an EDTA tube. Due to scheduling conflicts two samples were stored at 4C or -80C for no longer than 48 hours, all other samples were processed immediately after collection.

Sample preparation
The preparation of a sequencing library for distribution over 48 flow cells requires a substantial yield of high quality, high molecular weight genomic DNA. We optimized DNA extraction and library preparation for yield, quality, and speed using a modified Puregene (Qiagen, Hilden, Germany) genomic DNA extraction protocol, from the limited volume of blood. DNA was fragmented using g-TUBE (Covaris, Massachusetts, USA) and sequencing libraries were prepared with an SQK-LSK109 (Oxford Nanopore, Oxford, UK) protocol modified such that eight libraries were sufficient to load 48 PromethION flow cells. For barcoded samples, an additional EXP-NBD104 kit was used for barcoding. A detailed description of these methods can be found at dx.doi.org/10.17504/protocols.io.bvijn4cn.

Sequencing
While one PromethION flow cell can generate 200 gigabases (Gb) of data in 2-3 days of sequencing, our goal was to complete sequencing as quickly as possible. To this end, we distributed a DNA library from a single patient equally over 48 PromethION flow cells (Oxford Nanopore, Oxford, UK) and sequenced until we achieved a target range of greater than 170 Gb of data. After method optimization, we were able to generate as much as 200 Gb of data in as little as 1 hour 50 minutes. We dramatically reduced the sequencing cost-per-sample by washing the flow cells after sequencing and reusing the flow cells for multiple samples.

Base Calling and Alignment
Sequencing from 48 flow cells can yield more than 100 Gb per hour. While the PromethION tower is capable of locally base calling and aligning the data generated during sequencing, it cannot keep up with such a high rate of data generation. This increases the overhead compute time — base calling and alignment runtime beyond completion of sequencing — to almost 18 hours (220 Gb output).[1] To mitigate this, we implemented an approach that scaled multiple PromethION-tower like compute instances in a cloud computing environment (Google Cloud Platform) to achieve near real-time base calling and alignment of sequencing data. Going from a local to a cloud-based pipeline, we developed a framework where terabytes of data were transferred to the cloud storage in real-time during sequencing and distributed across the instances so as to minimize the tail latency. Specifically, as soon as sequencing begins, a script starts to periodically upload raw signal output files from the local tower to the cloud storage.

Simultaneously, multiple compute instances fetch batches of signal files designated to the particular instance from the cloud storage to perform base calling (Guppy v4.2.2) and alignment (Minimap2 v2.17-r974[2]) concurrently. The reads were aligned to the GRCh37 human genome reference genome.

Variant Calling

After alignment, we used a haplotype-aware long-read optimized variant caller (PEPPER-Margin-DeepVariant[3]) to identify single nucleotide variants (SNVs) and small insertions and deletions (indels). The pipeline included the structural variant caller Sniffles.[4] The pipeline is parallelized over multiple compute instances, each processing a specific set of contigs in order to rapidly generate an integrated variant call file. Across the samples, we achieved a mean heterozygous to homozygous and transitions to transversion ratio ratios were 1.5 and 2.0 respectively (**Figure S6 and Figure S7**)

Annotation

Small variants in homopolymer regions were annotated for use in the prioritization scheme. Structural variants were annotated with the frequency of similar variants in public and in-house SV catalogs derived from both short-read and long-read sequencing studies. Rare SVs that overlapped coding sequences were ranked based on the impacted gene and non-coding rare SVs were placed in a second tier of variants for curation.

Variant filtering and prioritization

Patient-specific gene target lists were developed for each case, in collaboration with ordering clinicians. While variant prioritization was not limited to the gene list, it ensured that genes highest on the clinical differential were examined rigorously. Alissa Interpret software (Agilent, Palo Alto, CA) was used to filter and prioritize likely deleterious variants, using a custom classification tree. Analysis was limited to variants that had gnomAD minor allele frequency<0.5% (in all non-bottleneck subpopulations) and that were in genes either associated with disease in the Online Mendelian Inheritance in Man catalog (“OMIM morbid”) or included on the target list. Variants not on the target gene list were further limited to those either within protein coding exons (including +/-2 intronic bases), or those with high confidence annotations in both the Human Gene Mutation Database (HGMD) and ClinVar.

Variants were then prioritized for review based on: (i) annotation in ClinVar and/or HGMD, (ii) inclusion on the target gene list, (iii) patient zygosity consistent with OMIM-annotated inheritance for associated disease, and (iv) predicted deleteriousness (e.g., truncating or missense variants in a gene highly intolerant of missense variation) (**see Figure S1 for more detail**). The prioritization scheme allowed rapid review and interpretation of likely actionable variants by setting a threshold of 4 or higher for review, and elevating the highest scoring variants for focused interpretation.

Variant curation and molecular board review

Variant interpretation was performed by one genetic counselor (MEG), one genomic scientist (DGF) and one American Board of Medical Genetics and Genomics board certified molecular geneticist (ES). This team worked in parallel, dynamically dividing and discussing variants as needed, and reviewed variants in order from the highest score (max=12) to the lowest (min=4).

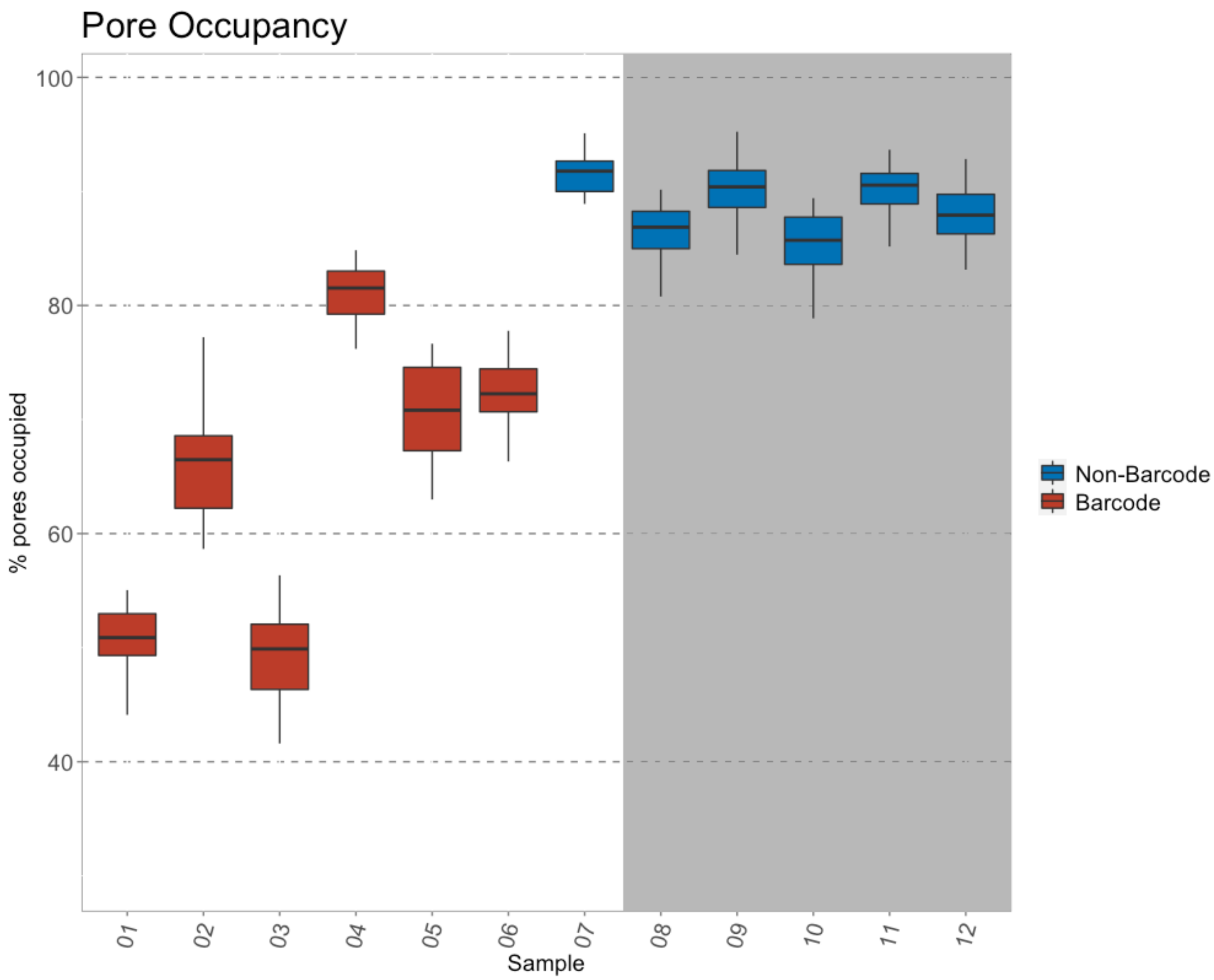
If a likely pathogenic, pathogenic, or otherwise suspicious variant was identified, it was reviewed in detail on a conference call by the clinical genetics and care team. If a consensus was reached that a pathogenic or likely pathogenic variant consistent with the patient’s phenotype was found, clinical confirmatory testing was completed either via single-site Sanger sequencing in Stanford's CLIA-CAP Clinical Genomics lab, and/or via clinical panel sequencing at an external CLIA/CAP laboratory.

Cost

We estimated the costs of our approach including DNA extraction, library preparation, sequencing, and computation and found those costs to range from $4971 - $7318 (for details of cost modeling, see **Table S6**). This is broadly comparable with similar rapid approaches previously reported using short read sequencing[5]. Given the daily cost of critical care is more than $10,000, rapid genome sequencing diagnostics well below this figure have been shown to be significantly cost saving and, as a result, are reimbursed by several payers.

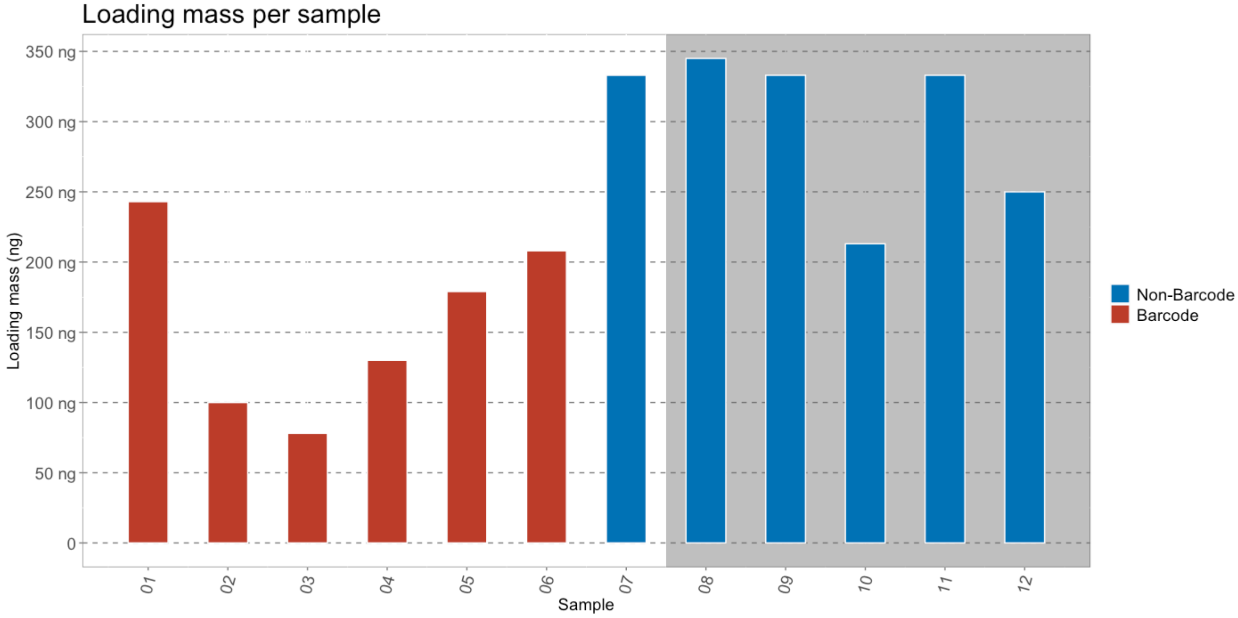
Supplementary Figures

Figure S1. Pore occupancy improvement for barcoded v/s non-barcoded samples



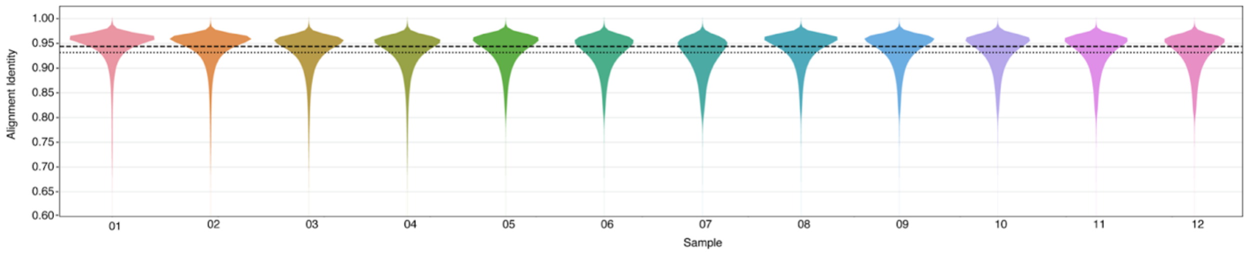
Eliminating barcoding results in higher pore occupancy. Barcoded libraries (red) have an average pore occupancy of 64% compared to non-barcoded libraries (Blue) of 82%. This difference was observed in phase one (white background), and it was decided to continue with the non-barcoding protocol in phase 2 (gray background). This increased pore occupancy is suspected to be a direct result of the increased yield of library in the non-barcoded samples, as seen in S1.

Figure S2. Loading mass improvement for barcoded v/s non-barcoded samples



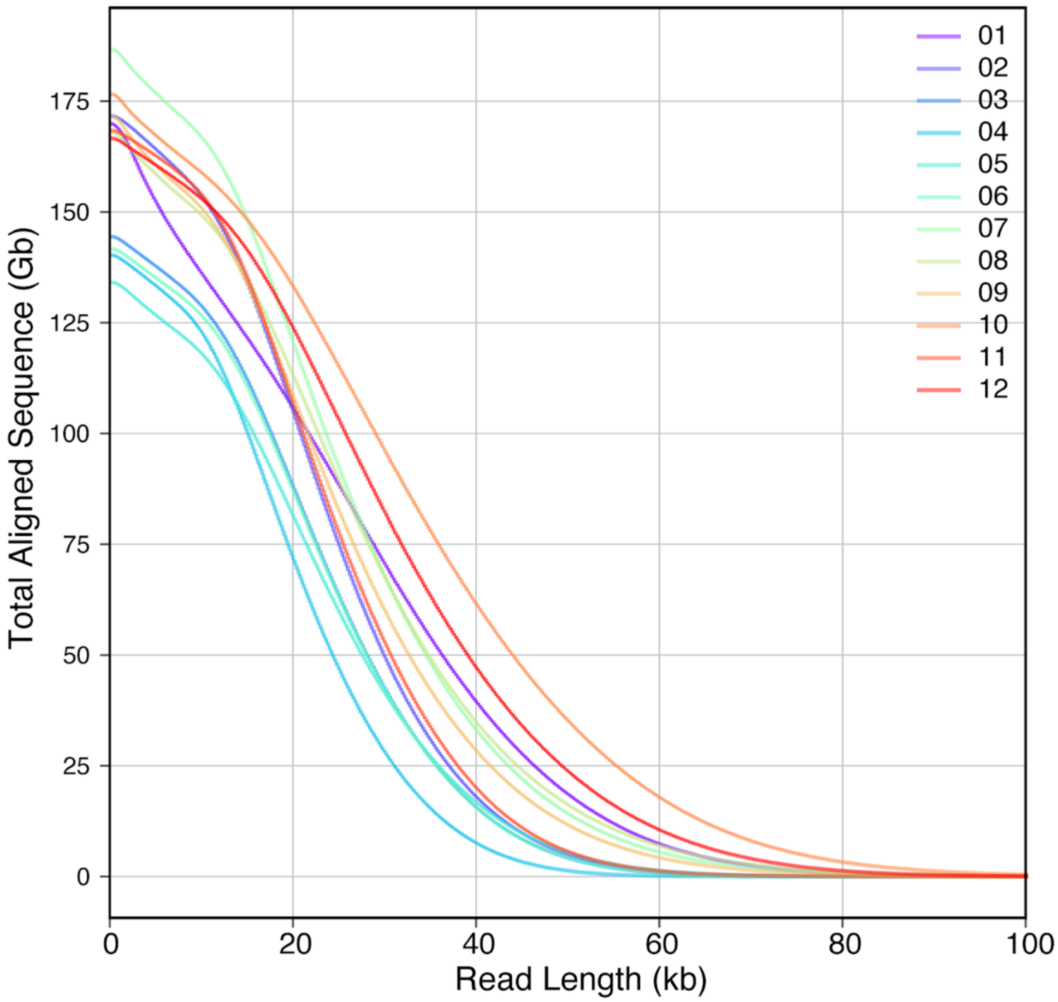
Ligating barcodes reduces the final yield of the sequencing library. The mean DNA loading mass per flow cell of barcoded libraries (red) was 155 ng (78 ng—243 ng) compared to non-barcoded libraries (blue) 333 ng (208 ng—345 ng). This difference was observed in phase one (white background), and it was decided to continue with the non-barcoding protocol in phase 2 (gray background).

Figure S3. Read alignment identity



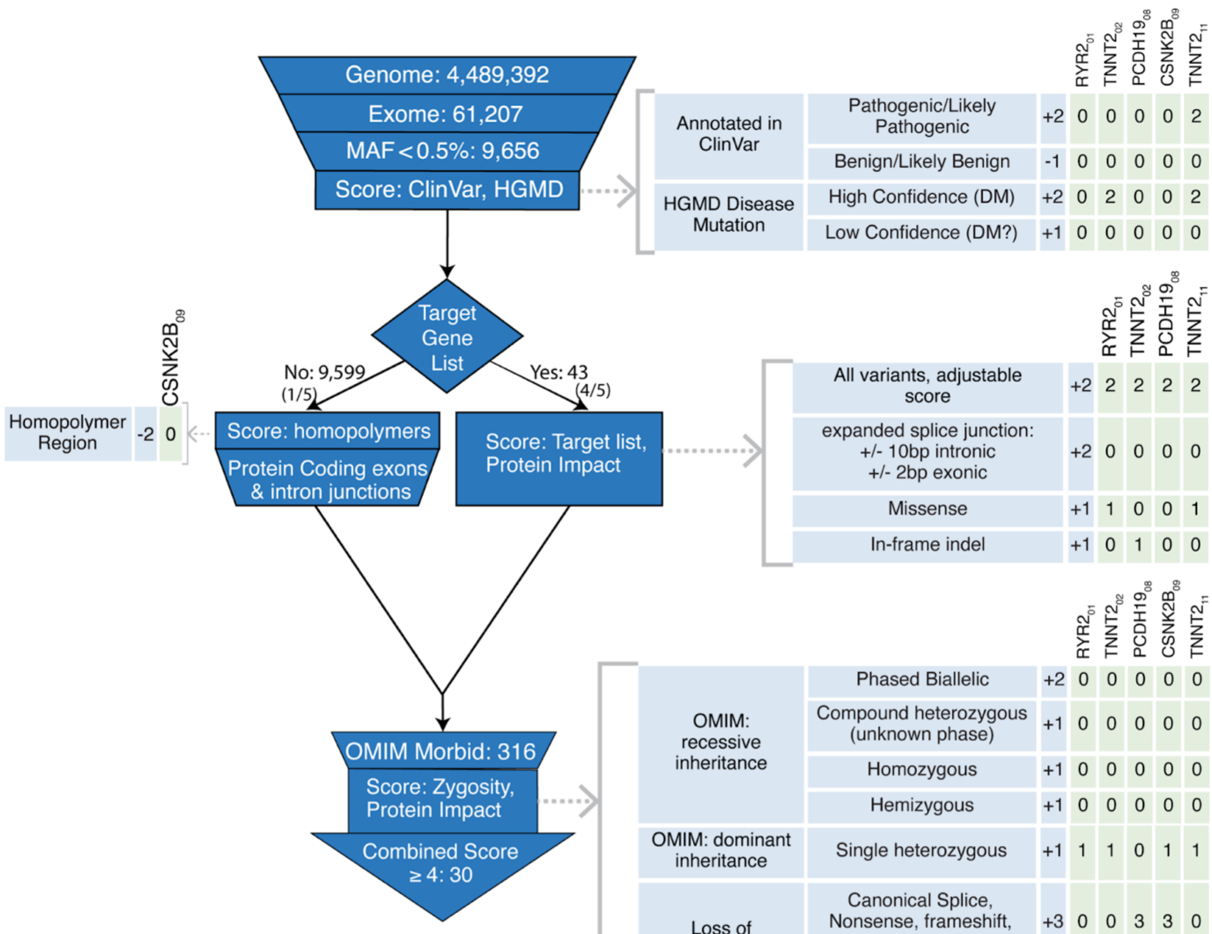
Alignment identities against GRCh37. Median of 0.944 is shown by the dashed line and the mean of 0.931 is shown by the dotted line.

Figure S4. NGx plot



Autosomal coverage for the samples as a function of the read length.

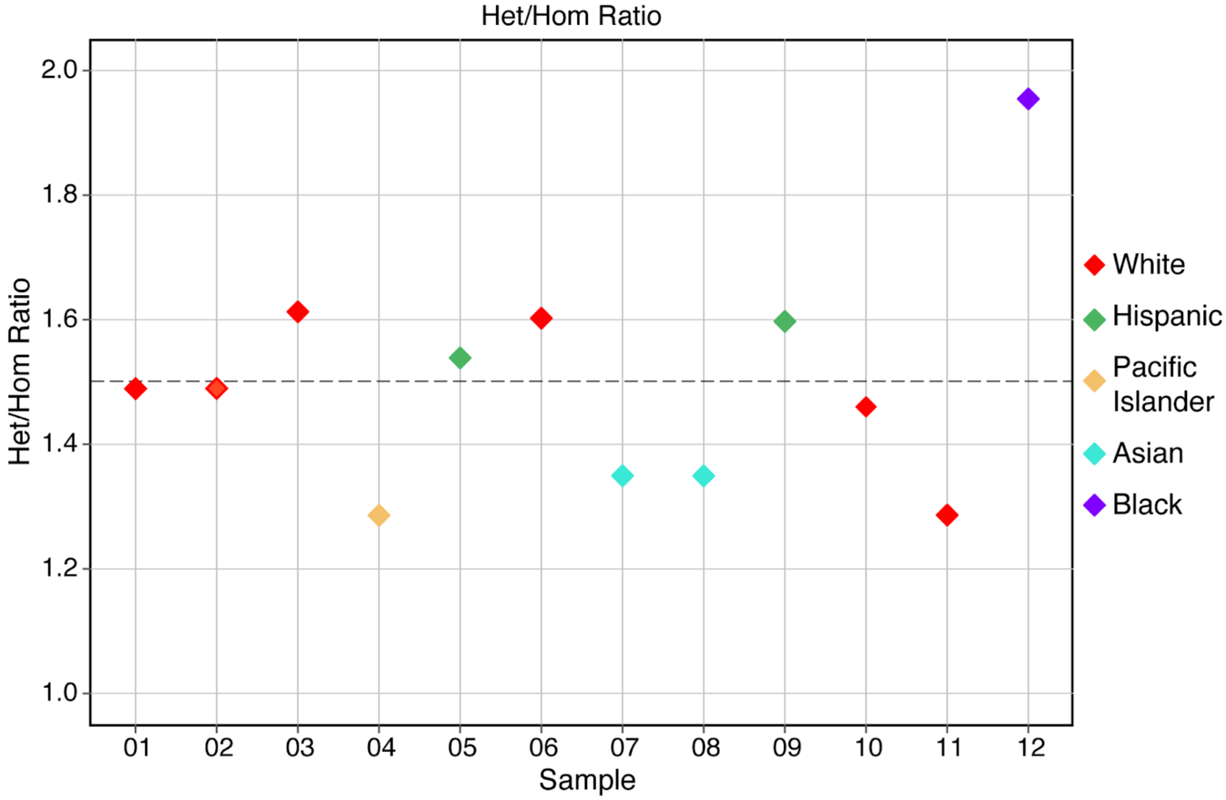
Figure S5. Variant filtration and prioritization



Patient ID	Gene	Final Score
01	RYR2	7
02	TNNT2	6
08	PCDH19	6
09	CSNK2B	5
11	TNNT2	8

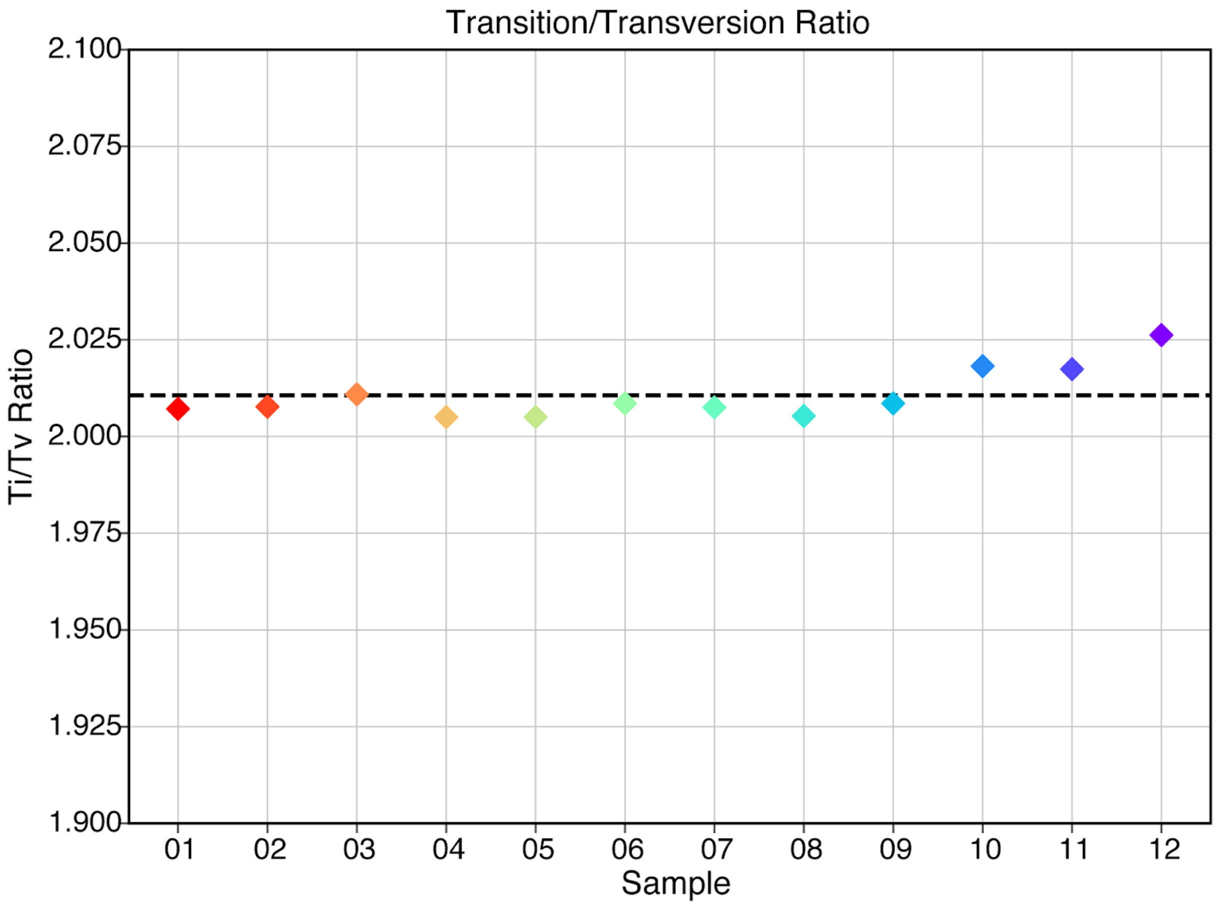
Variants are filtered and prioritized through a custom decision tree designed to surface the most likely pathogenic variants. Major filtration steps are depicted in dark blue (numbers represent average number of variants across all samples). Locations of prioritization scoring within the decision tree, as well as possible points assigned for variants meeting each criterion, are listed in light blue; points applied for each reported pathogenic/likely pathogenic variant are listed in adjacent green columns. The final prioritization score of each variant is shown in the dark green table.

Figure S6. Heterozygous/homozygous plot



Heterozygous/Homozygous ratio. Diamonds are colored based on the patient's ethnicity. The dashed line indicates a mean of 1.5.

Figure S7. Transition/transversion Ratio



Transition/Transversion ratio for the SNP calls across all the patient samples. The dashed line indicates a mean of 2.0

Supplementary Tables

Table S1. Demographic comparison

	US Census Data		Undiagnosed Diseases Network (N=2168)	Ultra-Rapid Whole Genome cohort (N=12)
	United States	California		
Sex				
Female	51.3%	50.7%	49.5%	41.7%
Male	48.7%	49.3%	50.5%	58.3%
Race/Ethnicity				
White	60.6%	37.0%	69.1%	50.0%
Hispanic/Latino	18.1%	39.1%	14.3%	16.7%
Asian	5.5%	14.4%	7.1%	16.7%
Pacific Islander	0.2%	0.3%	0.2%	8.3%
Black	12.3%	5.5%	5.9%	8.3%

Demographic Comparison. Self-reported demographic information from patients enrolled in the Ultra-Rapid Whole Genome study is representative of the general US and California population and of the Undiagnosed Diseases Network, a large cohort of patients with undiagnosed diseases suspected to have a genetic etiology.[6,7]

Table S2. Patient demographics, clinical presentation, and variant calls

Patient					Variants Calls/Prioritization		
ID	**Age**	**Sex**	**Clinical Presentation**	**Ethnicity**	**SNVS/INDELS**	**Prioritized 4+**	**Prioritized SV**
01	2 years	M	Cardiac Arrest, Ventricular Fibrillation	White	4,419,773	39	22
02	13 years	M	Dilated Cardiomyopathy, NSVT	White	4,442,280	28	20
03	6 months	F	Global developmental delay, Hypotonia, Microcephaly, Failure to thrive	White	4,478,350	35	17
04	5 years	M	Global developmental delay, Laryngeal cleft soft palate, Complex congenital heart disease, Choreiform movements, Ataxia	Pacific Islander	4,467,180	30	18
05	3 months	M	Recurrent Seizures, Focal seizures with R temporal focus, Thin corpus callosum, Apneic events with cyanosis	Hispanic	4,592,381	53	11
06	4 months	F	Cardiac Arrest	White	4,500,293	27	37
07	8 months	F	Cardiac Arrest, Torsades and polymorphic VT	Asian	4,482,314	37	25
08	6 months	F	Very frequent seizures in clusters after 6-month vaccinations	Asian	4,503,667	29	20
09	3 months	F	Seizure/Status Epilepticus, EEG: abundant multifocal sharps (biposterior)	Hispanic	4,619,267	28	35
10	2 weeks	M	Hypotonia, Pectus excavatum, Micrognathia	White	4,364,225	16	17
11	57 years	M	Left ventricular asymmetrical hypertrophy, LV/EF 40%	White	4,315,548	22	16
12	15 years	M	Dilated Cardiomyopathy, NSVT, LV/EF 15%	Black	4,770,449	20	22

The prioritization scheme places variants on a scale of 4–12. SV—structural variant; SNV—single nucleotide variant;indel—insertion or deletion; NSVT—non sustained ventricular tachycardia

Table S3. Clinical presentation and disease causing variants

Patient				Disease associated Variant							
ID	**Age**	**Sex**	**Presentation**	**Gene**	**Location**	**Details**	**Zygosity**	**ACMG Classification (Criteria)**	**Confirmation Testing**	**Diagnosis**	**Clinical management impacted**
01	2 years	M	Cardiac Arrest, Ventricular Fibrillation	*RYR2*	c.11621 C>T p.T3874I	Missense de novo	Heterozygous	Likely Pathogenic (PS2, PM2, PP3)	MGP	Catecholaminergic Polymorphic Ventricular Tachycardia	Sympathectomy, family screening
02	13 years	M	Dilated Cardiomyopathy, NSVT	*TNNT2*	c.487_489dup p.E163dup	INDEL (STR duplication) de novo	Heterozygous	Likely Pathogenic (PS4, PM2, PM6)	MGP, Sanger	Dilated Cardiomyopathy	Heart Transplant (irreversible cause), family screening
08	6 months	F	Very frequent seizures in clusters after 6- month vaccinations	*PCDH19*	c.2728G>T p.E910*	Predicted truncating de novo	Heterozygous	Likely Pathogenic (PVS1, PM6, PM2)	Sanger	*PCDH19*-related epilepsy	Selection of anti-seizure medications, access to ongoing clinical trials, counseling regarding prognosis and family planning
09	3 months	F	Infantile onset multifocal epilepsy with myoclonic seizures, status epilepticus	*CSNK2B*	c.73-1G>A	Splice Variant de novo	Heterozygous	Likely Pathogenic (PVS1, PS2, PM2)	MGP, Sanger	*CSNK2B*-related disorder/ Poirier-Bienvenu neurodevelopmental syndrome	Counseling regarding prognosis and family planning, avoided further extensive work-up
11	57 years	M	Left ventricular asymmetrical hypertrophy LV/EF 40%	*TNNT2*	c. 341C>T p.A104V	Missense	Heterozygous	VUS/ Likely Pathogenic (PM2)	Sanger	Hypertrophic Cardiomyopathy	Cardiac biopsy avoided, heart transplant (irreversible cause), family screening

Patient description and genetic findings. MGP--Multiple Gene Panel, INDEL--insertion/deletion, NSVT--non sustained ventricular tachycardia, STR--short tandem repeat, LV--left ventricular, EF--ejection fraction, VUS--variant of unknown significance. Transcripts: *RYR2* NM_001035.2, *TNNT2* NM_001001430.2 (ID02), *PCDH19* NM_001184880.1, *CSNK2B* NM_001320.6, *TNNT2* NM_001001430.1 (ID11)

Table S4. Summary statistics for sample preparation

Sample	01	02	03	04	05	06	07	08	09	10	11	12
Total DNA extracted (ug)	50.4	60.8	60	24	44	46.6	36	62	50	42	28	32
Total Input DNA to library prep (ug)	48	32	32	24	32	32	32	32	32	32	28	32
Total library recovery (ug)	11.7	4.8	3.76	5.2	8.6	10	16	16.6	16	10.24	16	12
% library recovered	24.38	15	11.75	21.67	26.88	31.25	50	51.88	50	32	57.14	37.5
Sequencing time (min)	180	162	260	217	225	346	489	110	122	144	136	162
Number of flow cells	48	48	48	40	48	48	48	48	48	48	48	48
Loading mass/flow cell (ng)	243	100	78	130	179	208	333	345	333	213	333	250
Number of Pores Available	329945	267234	186032	115732	138547	85762	51413	328017	297777	243207	207746	189072
Number of Channels Available	125170	114450	87847	57443	64490	43664	28775	122756	113875	98308	87366	82863
Mean Pore Occupancy (%)	49.7	65	49.5	80.3	68.8	72	89.5	85	89.2	84.1	89.6	86.4

Sample preparation and flow cell usage statistics. Quantification of DNA (inclusive of sequencing library) was accomplished using fluorometry (Qubit, Invitrogen) and represented in either micrograms (ug) or nanograms (ng). Number of available sequencing pores and channels was based on the primary sequencing mux scan. Mean pore occupancy was calculated by determining the average number of pores sequencing during the run divided by the total number of available pores. Barcoding was discontinued starting from patient 08.

Table S5. Stage-wise runtime across samples

Sample	01	02	03	04	05	06	07	08	09	10	11	12
DNA extraction (hr)	0.85	1.03	0.87	1.23	0.87	0.78	0.88	0.83	0.72	0.87	0.83	0.8
DNA fragmentation (hr)	0.63	0.32	0.22	0.27	0.3	0.22	0.2	0.22	0.2	0.18	0.2	0.2
Library Preparation (hr)	3.05	2.85	2.5	2.13	2.42	2.25	1.63	1.55	1.57	1.52	1.5	1.43
Sequencing (hr)	3	2.7	4.33	3.62	3.75	5.77	8.15	1.83	2.03	2.4	2.27	2.7
Basecalling, Alignment (hr)	10.35	4.2	4.78	4.28	4.25	6.18	8.63	2.28	2.65	2.8	2.62	3.12
Variant Calling (hr)	1.35	2.1	2.25	1.18	1	0.92	0.83	1.2	1.45	1.35	1.77	1.68
Idle time (hr)	2	0	0	0	0	0	0	0	0	0	0	0
Curation (hr)	0.45	0.77	3.13	2.22	4.75	2.67	3.22	1.33	1.58	1.28	0.38	1.73
Total time (hr)	18.68	11.27	13.75	11.31	13.59	13.02	15.39	7.41	8.17	8	7.3	8.96

Time expenditure for each stage of the rapid sequencing process. These measurements represent continuous time inclusive of time required for transition from one step to the next. Start and stop time was recorded when the DNA extraction protocol began, and when a definitive result was obtained respectively.

Table S6. Summary statistics for sequenced data

Sample	01	02	03	04	05	06	07	08	09	10	11	12
Number of Reads	18197600	14627670	13234243	14276670	13351924	13810830	17500416	14326476	16134470	14126585	13051305	11959458
Number of Passed Reads	17333484	13478052	11982411	12662292	11890482	12020842	15033695	13157707	14869334	12733570	11613172	10537576
% Passed Reads	95.3	92.1	90.5	88.7	89.1	87	85.9	91.8	92.2	90.1	89	88.1
Total Bases (Gb)	193	205	180	175	173	187	236	199	207	212	202	202
Passed Bases (Gb)	188	195	169	162	160	169	212	190	197	199	186	185
% Passed Bases	97.4	95.1	93.9	92.6	92.5	90.4	89.8	95.5	95.2	93.9	92.1	88.1
Mean QScore	12.5	12	11.5	11.2	11.5	11.1	10.7	11.9	11.8	11.5	11.4	11.2
Median QScore	13.1	12.7	12.2	11.9	12.2	11.7	11.3	12.6	12.4	12.1	12	11.9
N50	26.182	22.97	22.966	20.241	22.992	22.934	24.868	26.471	24.68	32.646	23.859	29.922

Sequencing data statistics. These data represent the quantity, quality and read length (N50) of the sequencing data generated by the PromethION flow cells. Passed reads and bases are required to reach a quality score (QScore) threshold of 7.

Table S7. Cost estimate

	Unit Price ($)	Quantity/run	Cost ($)
Sample Preparation			
Qiagen PureGene Blood Kit	4.5	8	36
Covaris G-Tube	35	8	280
NEB Reagents	38	8	304
LSK-109 (110) Kit	80	8	640
Flow cell wash Kit	10	48	480
ONT Expansion Packs	10	48	480
Miscellaneous Consumables	50	1	50
Sample Preparation per patient			**2270**
Computation			
Basecalling, Alignment Compute	8.818	45	397
Small Variant calling Compute	7.335	22	161
SV calling Compute	3.247	3	10
Computation per patient			**568**

PromethION 48 flowcells cost per patient		
	7 samples per week (360/yr)	**1 sample per week (54/yr)**
Cost per flowcell / **Re-uses per flowcell**	800 (for 18 re-uses) 680 (for 12 re-uses)	1120
18	2133	2987
12	2720	4480

Total cost per patient		
	7 samples per week (360/yr)	**1 sample per week (54/yr)**
Cost per flowcell / **Re-uses per flowcell**	800 (for 18 re-uses) 680 (for 12 re-uses)	1120
18	4971	5825
12	5558	7318

Cost Estimate for Sample Preparation, Sequencing and Computational resources. The sample preparation cost includes the total cost to prepare libraries across all 48 flow cells. The computational cost unit price is the estimated per hour cost provided by Google Cloud Platform and the quantity per run is based on the average number of hours estimated from our samples. The cost of flow cells has been calculated as a range where an average of 7 samples are sequenced per week and the highest is when only 1 sample is sequenced on an average per week. We have also provided a range of values based on the number of patients sequenced per set of 48 flow cells that can range between 12 and 18.

Acknowledgements

We would like to thank Oxford Nanopore Technologies for contributing to the cost of sequencing and reagents, Google for contributing to the cost of cloud computing, and NVIDIA for contributing to the cost of Parabricks. We would like to specifically thank George Vacek (NVIDIA) and Johnny Israeli (NVIDIA), Tom King (Google), Hermine Hovhannisyan (Google) for their contributions and guidance, Karl Kornel (Stanford) and the Stanford Research Computing Center for advising on the network and bandwidth requirements needed to complete this project, Terra Coakley, Brooke Gazzoli, and Brianna Mau Tucker for organizing the patient workbook, and Shirley C. Sutton for managing product orders.

We would like to thank all of the clinicians who referred patients to our study.

We would like to thank the patients and families who enrolled in our study and for their continued support to progress science and medicine.

Further Disclosures

Google employees did not have access to any patient data.

References

1. Sneha D. Goenka*, John E. Gorzynski*, Kishwar Shafin*, Dianna G. Fisk, Trevor Pesout , Jean Monlong , Tanner D. Jensen , Pi-Chuan Chang, Gunjan Baid, Jonathan A Bernstein, Jeffrey Christle , Karen P. Dalton, Daniel R. Garalde, Megan E. Grove, Joseph Guillory, Alexey Kolesnikov, Maria Nattestad, Maura R.Z. Ruzhnikov, Mehrzad Samadi, Ankit Sethia, Elizabeth Spiteri, Chris Wright, Katherine Xiong, Tong Zhu, Miten Jain, Fritz J. Sedlazeck, Andrew Carroll, Benedict Paten, Euan A. Ashley. Technical development of rapid whole genome nanopore sequencing and variant identification pipeline. In Press, Nature Biotechnology

2. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 2018;34(18):3094–100.

3. Shafin K, Pesout T, Chang P-C, et al. Haplotype-aware variant calling enables high accuracy in nanopore long-reads using deep neural networks [Internet]. bioRxiv. 2021 [cited 2021 May 15];2021.03.04.433952. Available from: https://www.biorxiv.org/content/10.1101/2021.03.04.433952v1

4. Sedlazeck FJ, Rescheneder P, Smolka M, et al. Accurate detection of complex structural variations using single-molecule sequencing. Nat Methods 2018;15(6):461–8.

5. Farnaes L, Hildreth A, Sweeney NM, et al. Rapid whole-genome sequencing decreases infant morbidity and cost of hospitalization [Internet]. npj Genomic Medicine. 2018;3(1). Available from: http://dx.doi.org/10.1038/s41525-018-0049-4

6. Explore Census Data [Internet]. [cited 2021 Nov 30];Available from: https://data.census.gov/cedsci/table?q=United%20States&g=0100000US_0400000US06&tid=ACSDP1Y2017.DP05&hidePreview=true

7. Gahl WA, Wise AL, Ashley EA. The Undiagnosed Diseases Network of the National Institutes of Health: A National Extension. JAMA 2015;314(17):1797–8.